

AD 633740

THE NEGATIVE BINOMIAL DISTRIBUTION:  
COMPUTATION OF THE MEDIAN AND THE MEAN ABSOLUTE DEVIATION

Definitions

variabl

discrete

,2,...,

is call

bution

when n

distrib

$f(x;n,p)$

numbered

precede

distrib

the geom

is the f

with its

Mean, Va

M

V

These va

From: J

CLEARINGHOUSE FOR FEDERAL SCIENTIFIC AND TECHNICAL INFORMATION			
Hardcopy	Microfiche		
\$ 1.00	\$ .50	8	pp
/ ARCHIVE COPY			

cal's distri-  
bution

s of this

of success,

trial

res will

of the

which has

,  $f(x;n,p)$

tribution

D D C

EP 23 1966

C  
ing function

J. Little, Inc.

AD 633740

THE NEGATIVE BINOMIAL DISTRIBUTION:  
COMPUTATION OF THE MEDIAN AND THE MEAN ABSOLUTE DEVIATION

Definitions

For  $n > 0$ ,  $0 < p < 1$ , and  $q = 1 - p$ , the distribution of the discrete variable  $x$ , having frequency function

$$f(x;n,p) = \binom{x+n-1}{x} p^n q^x = \binom{-n}{x} p^n (-q)^x, x = 0, 1, 2, \dots,$$

is called the negative binomial distribution. It also is called Pascal's distribution when  $n$  is a positive integer, and is called the geometric distribution when  $n = 1$ .

If  $n$  is a positive integer there are two well-known instances of this distribution. In a sequence of Bernoulli trials with probability  $p$  of success,  $f(x;n,p)$  is the probability that the  $n^{\text{th}}$  success will occur on the trial numbered  $(n+x)$ ; that is, it is the probability that exactly  $x$  failures will precede the  $n^{\text{th}}$  success. Also,  $f(x;n,p)$  is the frequency function of the distribution of the sum of  $n$  random, independent variables, each of which has the geometric distribution with frequency function  $f(x;1,p)$ ; that is,  $f(x;n,p)$  is the frequency function of the  $n^{\text{th}}$  convolution of the geometric distribution with itself, which will be evident from the generating function. D D C

Mean, Variance, and Higher Moments

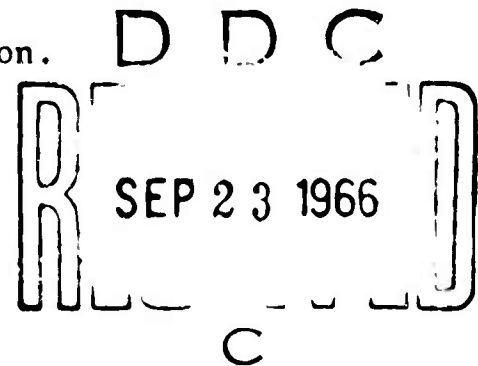
$$\text{Mean, } \mu = n q/p$$

$$\text{Variance from the mean, } \sigma^2 = n q/p^2$$

These values can be obtained by direct summation or from the generating function

From: James M. Dobbie

Revised by Little, Inc.



$$F(s) = \sum_{x=0}^{\infty} f(x;n,p) s^x = \left( \frac{p}{1-qs} \right)^n = p^n (1-qs)^{-n}$$

Higher moments can be computed from the formula

$$\sum_{x=k}^{\infty} \binom{x}{k} f(x;n,p) = \binom{n+k-1}{k} \left( \frac{q}{p} \right)^k ,$$

obtained by taking the  $k^{\text{th}}$  derivative of  $F(s)$  and putting  $s = 1$ . Such moments are needed in fitting polynomials in exponential smoothing. Thus, if

$$p_{T+t} = \sum_{(k)} a_k(T) t^k ,$$

the  $n^{\text{th}}$  exponentially-weighted average of the values of the polynomial for  $t \leq 0$  is

$$\begin{aligned} S_T^n(p) &= \alpha^n \sum_{x=0}^{\infty} \beta^x \binom{x+n-1}{n-1} p_{T-x} \\ &= \sum_{(k)} (-1)^k a_k(T) \sum_{x=0}^{\infty} x^k f(x;n,\alpha) \end{aligned}$$

However, if the polynomial is written in the form

$$p_{T+t} = \sum_{(k)} b_k(T) \binom{t+k-1}{k} ,$$

the  $n^{\text{th}}$  average is

$$\begin{aligned} S_T^n(p) &= \sum_{(k)} (-1)^k b_k(T) \sum_{x=k}^{\infty} \binom{x}{k} f(x;n,\alpha) \\ &= \sum_{(k)} (-1)^k \binom{n+k-1}{k} \left( \frac{\beta}{\alpha} \right)^k b_k(T) . \end{aligned}$$

It is for this reason that the second form of the polynomial is the preferred form.

### Median

In general, the equation

$$\sum_{x=0}^m f(x;n,p) = 1/2 ,$$

has no integral solution  $m$ . However, just as in the positive binomial distribution, the partial sum can be written in terms of the Incomplete Beta-function, which then can be used as the definition of the partial sum for non-integral values of  $m$ . In this way we can find a value of  $m$ , not necessarily an integer, that satisfies the equation. This value of  $m$  will be called the median.

The formula for the positive binomial distribution is

$$\sum_{x=0}^m \binom{n}{x} p^x q^{n-x} = I_q(n-m, m+1) = 1 - I_p(m+1, n-m), \quad (1)$$

where

$$I_p(a,b) = \frac{\int_0^p x^{a-1} (1-x)^{b-1} dx}{\int_0^1 x^{a-1} (1-x)^{b-1} dx}$$

is the Incomplete Beta-function. The formula appears in many books and can be proved easily by integration by parts.

The corresponding formula for the negative binomial distribution is

$$\sum_{x=0}^m \binom{x+n-1}{x} p^n q^x = I_p(n, m+1) = 1 - I_q(m+1, n). \quad (2)$$

From: James M. Dobbie

Arthur D. Little, Inc.

This formula is not readily available. It is stated, but not prominently, in the Introduction to Pearson's Tables of the Incomplete Beta-function and Pearson gives a proof (provided by Fieller) in *Biometrika*, Vol. XXV, pp. 160-161. A simpler proof is the following: Integrating by parts,

$$\begin{aligned} \int_0^p u^{n-1} (1-u)^m du &= \frac{1}{n} p^n q^m + \frac{m}{n} \int_0^p u^n (1-u)^{m-1} du \\ &= \frac{1}{n} p^n q^m + \frac{m}{n(n+1)} p^{n+1} q^{m-1} + \dots + \frac{m!}{n(n+1)\dots(n+m)} p^{n+m} \end{aligned}$$

Hence

$$I_p(n, m+1) = p^n \sum_{k=0}^m \binom{n+m}{k} p^{m-k} q^k \quad (3)$$

By induction on  $m$  it is easy to show that

$$\sum_{k=0}^m \binom{n+m}{k} p^{m-k} q^k = \sum_{k=0}^m \binom{x+n-1}{x} q^x. \quad (4)$$

Formula (2) is obtained from (3) and (4).

Although formula (2) has a meaning only when  $m$  is a non-negative integer, the integrals in the Incomplete Beta-function exist for non-integral values of  $m$ . We define the median of the negative binomial distribution to be the solution  $m$  of the equation

$$I_p(n, m+1) = 1/2 \quad (5)$$

A unique solution  $m \geq 0$  exists, provided  $p^n \leq 1/2$ .

### Mean Absolute Deviation

The mean absolute deviation from the median is

$$\Delta = \sum_{x=0}^{\infty} |x-m| f(x;n,p)$$

Let

$$\begin{aligned} [m] &= \text{integral part of } m \\ &= \text{largest integer that does not exceed } m. \end{aligned}$$

Then

$$\begin{aligned} \Delta &= \sum_{x=0}^{[m]} (m-x) f(x;n,p) + \sum_{x=[m]+1}^{\infty} (x-m) f(x;n,p) \\ &= m \left[ 2 \sum_{x=0}^{[m]} f(x;n,p) - 1 \right] + \mu - 2 \sum_{x=1}^{[m]} x f(x;n,p) \end{aligned}$$

Since

$$\begin{aligned} x f(x;n,p) &= \mu f(x-1;n+1,p), \\ \Delta &= m \left[ 2 I_p(n, [m]+1) - 1 \right] + \mu \left[ 1 - 2 I_p(n+1, [m]) \right] \end{aligned}$$

Other forms for  $\Delta$  can be obtained from

$$\begin{aligned} \sum_{x=1}^{[m]} f(x-1;n+1,p) &= \sum_{x=0}^{[m]} f(x;n,p) - \frac{1}{p} f([m], n+1, p) \\ &= \sum_{x=0}^{[m]} f(x;n,p) - \frac{([m]+1)}{\mu p} f([m]+1; n, p) \end{aligned}$$

From: James M. Dobbie

Arthur D. Little, Inc.

Two of these are

$$\Delta = (\mu - m) \left[ 1 - 2 I_p(n, [m] + 1) \right] + 2 \mu \binom{n+m}{n} p^n q^{[m]}$$

and

$$\Delta = (\mu - m) \left[ 1 - 2 I_p(n, [m] + 1) \right] + \frac{2}{p} \left( [m] + 1 \right) \left[ I_p(n, [m] + 2) - I_p(n, [m] + 1) \right] \quad (6)$$

The latter form is easy to use, since

$$[m] + 1 \leq m + 1 < [m] + 2 ;$$

that is, the two arguments involved are the two integers between which we interpolate in finding the solution of (5). Thus, to find  $\Delta$ , enter the tables of the Incomplete Beta-function and record the values

$$I_p(n, [m] + 1) \text{ and } I_p(n, [m] + 2)$$

for which

$$I_p(n, [m] + 1) \leq 1/2 \text{ and } I_p(n, [m] + 2) > 1/2$$

when the second argument has integral values. Interpolate to find  $m$  for which

$$I_p(n, m+1) = 1/2 \text{ and then substitute in (6).}$$

If  $m$  is an integer,

$$\Delta = \frac{(m+1)}{p} \left[ 2 I_p(n, m+2) - 1 \right] = 2n \binom{n+m}{m} p^{n-1} q^{m+1}$$

From: James M. Dobbie

Arthur D. Little, Inc.

If  $m$  is not an integer and we use linear interpolation between integral values to find it, then

$$\Delta = \left( \mu - m + \frac{1 + [m]}{p(m - [m])} \right) \left[ 1 - 2 I_p(n, [m] + 1) \right]$$

A quantity of interest in inventory problems is the expected amount by which  $x$  exceeds a given value  $k$ , that is, the expected back-order

$$B = \sum_{x=k}^{\infty} (x-k) f(x; n, p)$$

By the same arguments used to find  $\Delta$  we find

$$B = (\mu - k) \left[ 1 - I_p(n, k+1) \right] + \frac{(1+k)}{p} \left[ I_p(n, k+2) - I_p(n, k+1) \right]$$

for the negative binomial distribution.

### Examples

The values of  $\Delta$  and  $\sigma$  listed in the table below were computed primarily to test the hypothesis that

$$\Delta = k \sigma,$$

where  $k$  is 0.75 approximately. For  $p = 0.1$  it is necessary to use the formula

$$I_p(a, b) = 1 - I_q(b, a)$$

From: James M. Dobbie

Arthur D. Little, Inc.



For example,  $I_{0.9}(6,1) = 0.5314$  and  $I_{0.9}(7,1) = 0.4783$ ; from which  $I_{0.1}(1,6) = 0.4686$  and  $I_{0.1}(1,7) = 0.5217$ .

$p$	$n$	$\mu$	$\sigma$	$m$	$\mu-m$	$\Delta/\sigma$
0.1	1	9	9.5	5.6	3.4	0.69
	2	18	13.4	14.5	3.5	0.75
	3	27	16.4	23.4	3.6	0.76
	4	36	19.0	32.4	3.6	0.77
	5	45	21.2	41.4	3.6	0.77
0.5	1	1	0.71	0	1	0.71
	2	2	2.00	1	1	0.75
	3	3	2.45	2	1	0.76
	4	4	2.83	3	1	0.77
	5	5	3.16	4	1	0.78
0.9	10	1.11	1.11	0.43	0.68	0.88
	20	2.22	1.57	1.51	0.71	0.85
	30	3.33	1.92	2.63	0.70	0.83
	40	4.44	2.22	3.74	0.70	0.82
	50	5.56	2.48	4.85	0.71	0.81

From: James M. Dobbie/rc

Arthur D. Little, Inc.